Postal Regulatory Commission
Submitted 1/15/2015 3:49:22 PM
Filing ID: 91162
Accepted 1/15/2015

BEFORE THE
POSTAL REGULATORY COMMISSION
WASHINGTON, D.C. 20268–0001

PERIODIC REPORTING (PROPOSAL THIRTEEN)	Docket No. RM2015-7

RESPONSES OF THE UNITED STATES POSTAL SERVICE TO QUESTIONS 17-18 OF CHAIRMAN'S INFORMATION REQUEST NO. 1
(January 15, 2015)

The United States Postal Service hereby provides its responses to Questions 17-18 of Chairman's Information Request No. 1, issued January 6, 2015. The questions are stated verbatim and followed by the response. Responses to all other questions were filed on January 12, 2015

Respectfully submitted,

UNITED STATES POSTAL SERVICE

By its attorney:

Eric P. Koetting

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260-1137
(202) 277-6333
January 15, 2015

17. Please refer to the Report at 73, which states that the Postal Service removed "some of those higher order terms, including cross-products… that are not statistically significant," as a method of dealing with moderate multicollinearity.

a. Please discuss whether other methods of ameliorating the negative effects of multicollinearity were considered, and if so, why they were not employed. Please include a discussion of the following methods:
   - i. Centering variables around their respective means,
   - ii. Using a logarithmic transformation of variables,
   - iii. Variable Selection Methods such as:
     1. Sequential Regression
     2. Best Subset Selection
     3. Shrinkage Models
   - iv. Newey-West Robust Estimation (*See* Econometric Foundations, Mittelhammer, Judge, and Miller, at 392).

b. Did the Postal Service test for model improvement after omitting variables to ameliorate multicollinearity using a measure that could compare goodness of model fit with all variables and without excluded variables, such as the Information Criterion or other measures of model fit?

c. If the Postal Service did perform versions of any of the above-mentioned tests of this sort, please provide all data, programs, and associated logs necessary to reproduce the results from each test.

RESPONSE:

a. In choosing a method to ameliorate the negative effects of multicollinearity on an estimated model, one should consider both the degree and the possible sources of the problem, as well as the nature of the proposed remedy. In addition, one should consider the nature of the model being estimated and the data used for that estimation. For example, the data used for estimating the regular delivery time equation includes both an FSS mail variable and a sequence mail variable. This is noteworthy as many ZIP Code days have no FSS mail (if the ZIP Code does not have FSS equipment) and many ZIP Code days have no sequenced mail because no mailings of that type were received on a given day. This means that both variables can regularly take on a value of zero. The existence of these zero values precludes the use of a logarithmic transformation of the right-hand-side variables.

Moreover, the regular delivery time model is quadratic and thus has higher-order cross product terms that are themselves a source of multicollinearity. In that instance, it has been argued in the literature that mean-centering is ineffective as a multicollinearity treatment.[1] Next, Newey-West robust standard errors are sometimes used as the basis for determining which variables should be dropped from a regression equation because they correct for both autocorrelation and heteroscedasticity and thus provide better inferences. However, given the short time dimension of the regular delivery data, White robust standard errors were used for hypothesis testing instead of Newey-West standard errors, as the latter depends upon the existence of lagged time series data. In concept, nevertheless, the actual variable section method was based upon a procedure that used robust standard errors as is done when applying Newey-West standard errors.

In sum, given the moderate degree of multicollinearity, a reasonable variable selection method, based upon robust standard errors, was applied. This particular approach was advantageous because it was parsimonious but produced a model with all statistically significant coefficients. In addition, it did not

[1] For example see, Raj Echambadi and James D. Hess, "Mean-Centering Does not Alleviate Collinearity Problems in Moderated Multiple Regression Models," Marketing Science, Vol. 26, No.3, May-June 2007 at 438-445.

require elimination of a large number of terms and thus did not do violence to the flexibility inherent in the quadratic functional form.

b. Because individual t-tests were employed for each variable excluded, an overall goodness-of-fit test was not applied. One could, however, employ an F-test of the hypothesis that all of the dropped coefficients were jointly insignificant. That F-test has the following formula:

$$F_0 = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-(k+1))}$$

In that formula, SSR_r is the sum of squared residuals from the restricted model (with the six cross-product terms dropped), $SSRu_r$ is the sum of squared residuals from the unrestricted model (with the six cross-product terms included), n is the number of observations, k is the number of independent variables in the unrestricted model, and q is the number of variables dropped. This F-statistic is distributed with q and n-k degrees of freedom. The following table provides the results of the F-test indicating support of the null hypothesis that the omitted variables had zero coefficients, and were thus appropriately dropped.

Calculated Value	Critical Value	Conclusion
1.671	2.098	Do not reject the null hypothesis that the omitted coefficients are equal to zero.

c. Not applicable

18. Please refer to the Report, pages 84-85, where the Postal Service states that "[r]eview of the data for the 44 observations reveals nothing to suggest that the observations contain data errors or do not come from valid ZIP Codes that perform standard city carrier delivery operations… it is preferred to leave them in the data set when estimating the regression."

a. Did the Postal Service run a test to determine whether the F-values for the model with and without the 44 observations in question were statistically different?

b. Did the Postal Service run any goodness of fit tests to determine if there was a statistically significant difference in model fit? If not, please explain.

c. If the Postal Service did perform a version of the above-mentioned test, please provide all data, programs, and logs associated with necessary to reproduce the results from each test.

RESPONSE:

a. No. Given that there was no evidence that the 44 observations were outliers and given that the results (repeated below) of estimating the model with and without the 44 observations were virtually the same, there appeared to be no need for such a test..

Table 38
Delivery Time Variabilities from the Regular Delivery Equation

	All Observations	Dropping 44 Outliers	Difference
DPS Volume	0.1676	0.1594	-0.0083
Cased Volume	0.0699	0.0763	0.0064
Sequenced Volume	0.0338	0.0344	0.0006
FSS Volume	0.0295	0.0303	0.0008
Collection Volume	0.0541	0.0569	0.0029
Delivery Points	0.5491	0.5522	0.0030

b. No. Please see the answer to part a. above.

RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO CHAIRMAN'S INFORMATION REQUEST NO. 1

c. Not applicable However, below is the SAS output for estimation of the model with the 44 observations dropped.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: delivery_hrs

Number of Observations Read	3441
Number of Observations Used	3441

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	26	7157151	275275	886.81	<.0001
Error	3414	1059738	310.40954		
Corrected Total	3440	8216889			

Root MSE	17.61844	R-Square	0.871
Dependent Mean	93.12179	Adj R-Sq	0.87
Coeff Var	18.91979		

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent		
						Standard Error	t Value	Pr > |t|
Intercept	1	-16.67334	1.63151	-10.22	<.0001	1.29895	-12.84	<.0001
fssdum	1	3.98607	1.21693	3.28	0.0011	1.27744	3.12	0.0018
dps	1	0.00045628	0.00009738	4.69	<.0001	0.00010174	4.48	<.0001
dps2	1	-7.05E-09	8.48E-10	-8.31	<.0001	9.85E-10	-7.16	<.0001
cm	1	0.00107	0.00023466	4.57	<.0001	0.00024207	4.43	<.0001
cm2	1	-2.28E-08	5.23E-09	-4.36	<.0001	4.93E-09	-4.63	<.0001
seq	1	0.00092798	0.00008882	10.45	<.0001	0.00009759	9.51	<.0001
seq2	1	-1.98E-08	3.13E-09	-6.32	<.0001	3.18E-09	-6.22	<.0001
fss	1	0.00228	0.00031425	7.26	<.0001	0.00035496	6.43	<.0001
cv	1	0.00090674	0.00040947	2.21	0.0269	0.00045236	2	0.0451
cv2	1	-1.03E-07	1.91E-08	-5.37	<.0001	2.15E-08	-4.78	<.0001
pd	1	0.00668	0.00026818	24.9	<.0001	0.00026814	24.9	<.0001
pd2	1	-1.20E-07	1.27E-08	-9.46	<.0001	1.38E-08	-8.66	<.0001
dpscm	1	1.84E-08	3.94E-09	4.66	<.0001	4.56E-09	4.03	<.0001
dpscv	1	-4.01E-08	8.42E-09	-4.77	<.0001	9.94E-09	-4.03	<.0001
dpspd	1	4.00E-08	6.39E-09	6.27	<.0001	8.12E-09	4.93	<.0001
cmcv	1	9.24E-08	1.73E-08	5.34	<.0001	1.99E-08	4.65	<.0001
cmpd	1	-5.58E-08	1.41E-08	-3.96	<.0001	1.83E-08	-3.05	0.0023
fsscv	1	1.26E-07	2.22E-08	5.65	<.0001	2.42E-08	5.19	<.0001
fsspd	1	-1.02E-07	1.48E-08	-6.94	<.0001	1.93E-08	-5.32	<.0001
cvpd	1	1.29E-07	2.49E-08	5.19	<.0001	3.23E-08	3.99	<.0001
dt	1	43.68405	3.26705	13.37	<.0001	3.01696	14.48	<.0001
dt2	1	-26.15822	3.16547	-8.26	<.0001	3.09758	-8.44	<.0001
mpdp	1	76.95774	21.45229	3.59	0.0003	11.696	6.58	<.0001
mpdp2	1	-130.24016	47.54589	-2.74	0.0062	19.94342	-6.53	<.0001
busrat	1	-45.32677	11.11101	-4.08	<.0001	9.74638	-4.65	<.0001
busrat2	1	51.13818	18.69981	2.73	0.0063	14.36059	3.56	0.0004